SUPPLEMENTARY INFO

Isabel Nepomuceno [1], Francisco Azuaje [2], Yvan Devaux[2], Petr V. Nazarov[3], Arnaud Muller[3], Jesús S. Aguilar-Ruiz[4], Daniel R. Wagner [2,5]

[1]Dpto. Lenguajes y Sistemas Informáticos, Universidad de Sevilla,Spain
[2]Laboratory of Cardiovascular Research, CRP-Santé,Luxembourg
[3]Microarray Center, CRP-Santé, Luxembourg
[4]School of Engineering, Pablo de Olavide University, Spain
[5]Division of Cardiology, Centre Hospitalier, Luxembourg

1 BENCHMARK ANALYSIS

Table 1. Topological network parameters from benchmark dataset

Clinical outcome	*SATuRNo* AD	*SATuRNo* Non AD	COEXPRESS AD	COEXPRESS Non AD	PPC AD	PPC Non AD
Number of nodes	27	15	26	28	62	45
Number of edges	23	13	53	54	71	64
Connected components	5	2	9	10	27	18
Diameter	4	6	1	1	1	1
Density	0,06	0,12	0,16	0,14	0,03	0,06

AD is Alzheimer's Disease and non AD is control

Table 2. Intersection

	SATuRNo ⋂ COEXPRESS AD	*SATuRNo* ⋂ COEXPRESS non AD
genes	8	2
edges	4	1

AD is Alzheimer's Disease and non AD is control

Table 3. Intersection

	SATuRNo ⋂ PPC AD	*SATuRNo* ⋂ PPC non AD
genes	9	2
edges	0	0

AD is Alzheimer's Disease and non AD is control

Table 4. Intersection

	C OECPRESS ⋂ PPC AD	C OECPRESS ⋂ PPC non AD
genes	26	28
edges	0	0

2 SYSTEMS-BASED PROGNOSIS AFTER MI

We compared the predictive capability of our approach against standard classification models based on gene expression data. We trained several classifiers using the datasets under study. We trained "lazy learning" classifiers (Ib1, Ibk and Kstar), classifiers based on decision trees (ADTree, LMT, J48, NB Tree, Random Forest and Random Tree) and probabilistic classifiers based on Bayesian Statistics. Finally, we trained Support Vector Machine classifiers, which have been shown to be powerful and robust models in cancer and other research areas.

Although *SATuRNo* did not outperform all the classifiers, it showed one of the highest classification accuracy than those produced by the other methods, including Support Vector Machine classifiers. Only J48 classifier outperformed our approach using in its model the gene c20orf20 that has been recently suggested as a potential therapeutic target in colorectal cancer. The performance our approach suggests that networks estimated by our approach can provide the basis for relatively accurate classification models. These

results encouraged us to investigate our method as a new strategy to discover potential biomarkers of clinical outcome after Myocardial Infarction. This is specially motivated by the fact that our method, unlike traditional approaches, can provide mechanistic insights of processes and associations underlying the clinical conditions through specific network-based visualizations.

Table 5. Heart dataset: baseline comparison

IB1	56,25%
IBK	56,25%
Kstar	50%
AD Tree	65.62%
LMT	71.87%
J48	90.62%
NB Tree	62.5%
Random Forest	59.37%
Random Tree	40.62%
Naive Bayes	65.62%
SATuRNo	72,145%
Support Vector Machine	68.75%

The predictive capability of our approach against standard classification models based on gene expression data.

3 INGENUITY PATHWAY ANALYSIS

Ingenuity Pathway Analysis (IPA) is a pathway and function exploration tool that use prior knowledge. The genes from the networks were overlaid onto a global molecular network (GMN) developed from information contained in Ingenuity's Knowledge Base. After that, a functional analysis of those GMNs identified the biological functions and/or diseases that were most significant to the molecules in the network. From bad prognosis IPA reported the networks shown in Table 6. From good prognosis IPA reported networks in Table 7. Furthermore, this tool highlighted several genes as known biomarkers, see Tables 8 and 9.

Table 6. Heart dataset: IPA Analysis from bad prognosis

Top Biological Function	p-value
1. Cell Cycle, Hematological Systems Development and Function, Hematopoiesis	10^{-22}
2. Cell Death, Hematological Disease, Immunological Disease	10^{-19}
3. Cell-To-Cell Signaling and Interaction, Tissue Development, Hematological System Development and Function	10^{-4}
4. Cancer, Hematological Disease	10^{-3}
5. Behavior, Cardiac Edema, Cardiovascular Disease	10^{-3}

Associated network functions reported by Ingenuity Pathway Analysis (IPA) tool.

Table 7. Heart dataset: IPA Analysis from good prognosis

Top Biological Function	p-value
1. Behavior, Cardiac Edema, Card.Disease	10^{-3}
2. Cancer, Genetic Disorder, Reproductive Systems Disease	10^{-3}
3. Lipid Metabolism, Small Molecul Biochemistry, Molecular Transport	10^{-3} 10^{-3}
4. DNA Replication, Recombination, and Repair, Energy Product. and Nucleic Acid Metabolism	10^{-3}
5. Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair	10^{-2}

Associated network functions reported by Ingenuity Pathway Analysis (IPA) tool.

Table 8. Heart dataset: biomarkers from bad prognosis network

Gene	Disease
BCAM	Diagnosis of breast cancer
RPS4Y1	Prognosis of Ewing's sarcoma

Biomarkers presented in the bad prognosis network.

Table 9. Heart dataset: biomarkers from good prognosis network

Gene	Disease
DLK1	Bronchoalveolar adenocarcinoma in humans is associated with downregulation of human DLK1 (real-time RT-PCR)
HINT1	Prognosis of Ewing's sarcoma
RPS4Y1	Prognosis of Ewing's sarcoma
SARNP	diagnosis of bladder cancer

Biomarkers presented in the good prognosis network.

Table 10. Heart dataset: genes and descriptions

Gene Symbol	Description	Location	Type
ATP5SL	ATP5S-like	unknown	other
BCAM	basal cell adhesion molecule (Lutheran blood group)	Plasma Membrane	transmembrane receptor
COMMD9	COMM domain containing 9	unknown	other
CYB561D2	cytochrome b-561 domain containing 2	unknown	enzyme
DCTN1	dynactin 1 (p150, glued homolog, Drosophila)	Cytoplasm	other
EDF1	endothelial differentiation-related factor 1	Nucleus	transcription regulator
GTF3C4	general transcription factor IIIC, polypeptide 4, 90kDa	Nucleus	transcription regulator
SH3GLB2	SH3-domain GRB2-like endophilin B2	Cytoplasm	other
SMR3A	submaxillary gland androgen regulated protein 3A	Extracellular Space	other
VSTM1	V-set and transmembrane domain containing 1	unknown	other
ZNF511	zinc finger protein 511	Nucleus	other
APOF	apolipoprotein F	Extracellular Space	transporter
ARL6IP4	ADP-ribosylation-like factor 6 interacting protein 4	Nucleus	other
AUP1	ancient ubiquitous protein 1	Cytoplasm	other
BANF1	barrier to autointegration factor 1	Nucleus	other
CALN1	calneuron 1	unknown	other
CAND2	cullin-associated and neddylation-dissociated 2 (putative)	Nucleus	transcription regulator
CCNO	cyclin O	Nucleus	enzyme
SARNP	SAP domain containing ribonucleoprotein	Nucleus	other
CLEC4M	C-type lectin domain family 4, member M	Plasma Membrane	other
DLK1	delta-like 1 homolog (Drosophila)	Extracellular Space	other
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	Nucleus	transcription regulator
B9D1	B9 protein domain 1	Extracellular Space	other
FGF22	fibroblast growth factor 22	Extracellular Space	growth factor
FGFRL1	fibroblast growth factor receptor-like 1	Plasma Membrane	transmembrane receptor
GCN1L1	GCN1 general control of amino-acid synthesis 1-like 1 (yeast)	Cytoplasm	translation regulator
HERV-FRD	HERV-FRD provirus ancestral Env polyprotein	unknown	other
HINT1	histidine triad nucleotide binding protein 1	Nucleus	enzyme
HIST1H2AE	histone cluster 1, H2ae	Nucleus	other
IGK@	immunoglobulin kappa locus	Extracellular Space	other
KATNAL2	katanin p60 subunit A-like 2	unknown	other
LAMB4	laminin, beta 4	unknown	other
LY6K	lymphocyte antigen 6 complex, locus K	unknown	other
MEN1	multiple endocrine neoplasia I	Nucleus	transcription regulator
MPDU1	mannose-P-dolichol utilization defect 1	Cytoplasm	other
MRPS18A	mitochondrial ribosomal protein S18A	Cytoplasm	other
NDUFB8	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8, 19kDa	Cytoplasm	enzyme
OBFC2B	oligonucleotide/oligosaccharide-binding fold containing 2B	unknown	other
OR2T3	olfactory receptor, family 2, subfamily T, member 3	unknown	other
PLD3	phospholipase D family, member 3	Cytoplasm	enzyme
CENPV	centromere protein V	Nucleus	other
RAI1	retinoic acid induced 1	Cytoplasm	other
RPS4Y1	ribosomal protein S4, Y-linked 1	Cytoplasm	other
STAB2	stabilin 2	Plasma Membrane	transmembrane receptor
TBL1XR1	transducin (beta)-like 1 X-linked receptor 1	Nucleus	transcription regulator
TMEM106C	transmembrane protein 106C	unknown	other
ZNF542	zinc finger protein 542	unknown	other

Information relative to genes involved in the two networks (good and bad prognosis networks). Note: reference set is human transcriptome.